Roteiro 05 de Outubro de 2012

1. Acessar a base de paridade disponível em https://dl.dropbox.com/u/24941619/data_paridade.arff
2. Definir split = 50%
3. Aplicar o k-***nn {3,5,9}***
4. A acurácia melhorou ou piorou com a variação dos valores de K? Por que você acha que isso aconteceu?

5. Acesse o conjunto de dados *Wine Quality* a partir da página da UCI: http://archive.ics.uci.edu/ml/datasets/Wine+Quality
6. Salvar os arquivos winequality-red.csv e winequality-white.csv
7. Em cada arquivo, substituir todas as ocorrências de “;” por “,” usando algum editor de texto.
8. Fazer para os dois arquivos:
    1. Aplicar o k-***nn {13}*** e anotar o erro médio quadrático.
    2. Aplicar o filtro weka.attributeSelection.CfsSubsetEval.
    3. Anotar ao atributos que foram removidos
    4. Aplicar o k-***nn {13}*** e anotar o erro médio quadrático.
    5. Testar o k-***nn {13}*** para os atributos removidos. Anotar o erro médio quadrático.
    6. Em qual dos testes obteve-se o maior valor do erro médio quadrático? Discuta a influência da seleção de atributos nesse conjunto de dados.

9. Repetir o ítem 8, usando a rede neural MLP (weka.classifiers.functions.MultilayerPerceptron) em vez do knn.